Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 36 (2014) 278 – 284

Procedia Computer Science

Complex Adaptive Systems, Publication 4
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2014-Philadelphia, PA

On a Multidimensional Cluster Scaling

Mika Sato-Ilic[a], Peter Ilic[b]

[a]*Faculty of Systems and Information Engineering, University of Tsukuba, Tennodai 1-1-1, Tsukuba, Ibaraki 305-8573, Japan*
[b]*Faculty of Literature, University of Toyo, 5-28-20 Hakusan, Bunkyo-ku, Tokyo 112-8606, Japan*

Abstract

This paper proposes a multidimensional cluster scaling (MDCS) which obtains scaling of multidimensional clusters (or groups). Conventional multidimensional scaling is well known as a method to obtain the scaling of multidimensional objects. However, there is no method to obtain the scaling of the clusters. The merit of this scaling method is its applicability for the analysis of large amounts of data such as big data. Since the purpose of the scaling method is to obtain the latent structure of a given data in a lower dimensional space in order to summarize the data features and the visualization of the data structure, for large amounts of data, a loss of the data information through the reduction of the dimensions has been a main problem with the use of the scaling method. The proposed method can solve this problem by the use of clusters of objects. Several numerical examples show the better performance of the proposed method.

Peer-review under responsibility of scientific committee of Missouri University of Science and Technology
Keywords: labeled data; fuzzy clustering; cluster-scaled data transformation; multidimensional scaling

1. Introduction

Conventionally, the main purpose of data analyses is to summarize data using a smaller number of factors or the visualization in a lower dimensional space. However, when the amount of data is large, the reduction of many dimensions of the data space or the summarization using the much smaller number of factors cannot work efficiently due to the large loss of information of the original data. However, if the original data has additional information of groups, then the information might be used efficiently for capturing latent structure of the large amount of data. This paper proposes a method of MDS in which we use the additional information of data groups (labeled data sets) to the conventional MDS. Since the MDS uses the single scaled coordinate to show the visualization of the similarity structure among data, we need to transform the labeled data sets by using a single scaling. For this purpose, we use the classification structure as the scaling over the dissimilarity structures of labeled data sets in order to

*Corresponding author. Tel:+81-29-853-5006; fax:+81-29-853-5006
E-mail address: mika@risk.tsukuba.ac.jp

Peer-review under responsibility of scientific committee of Missouri University of Science and Technology
doi:10.1016/j.procs.2014.09.094

implement the capability to be comparable over the different labeled data sets. Using the transformed data, we propose a multidimensional scaling for clusters (or groups). We call this method the multidimensional cluster scaling (MDCS).

This paper consists of the following: Section 2 describes labeled data. Section 3 proposes a transformation method from a set of dissimilarity matrices of groups to single dissimilarity data based on a cluster-based scaling. Section 4 proposes a multidimensional cluster scaling (MDCS). Section 5 describes numerical examples and section 6 contains conclusions.

2. Labeled Data

Suppose X to be a given data matrix consisted of n objects and p variables as follows:

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_n \end{pmatrix}, \quad \boldsymbol{x}_i = (x_{i1}, \cdots, x_{ip}), \quad i = 1, \cdots, n. \tag{1}$$

In equation (1), a set of n objects $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ are labeled into T groups. We represent this status as follows:

$$\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\} = \{\boldsymbol{x}_1^{(1)}, \cdots, \boldsymbol{x}_{n_1}^{(1)}, \boldsymbol{x}_{n_1+1}^{(2)}, \cdots, \boldsymbol{x}_{n_1+n_2}^{(2)}, \cdots, \boldsymbol{x}_{n_1+\cdots+n_{T-1}+1}^{(T)}, \cdots, \boldsymbol{x}_n^{(T)}\}, \ n = n_1 + \cdots + n_T. \tag{2}$$

In equation (2), $\boldsymbol{x}_i^{(t)}, \ i = 1, \cdots, n, \ t = 1, \cdots, T$ shows an object i assigned to a group t. Group t has n_t objects.

3. Cluster-Scaled Data Transformation for Labeled Data

In equation (2), if objects in each group are the same, that is n_t are the same for all T groups, then we can treat this data as a 3-way data. There is a MDS for the 3-way data adapted scaling difference of the different groups. This method obtains a result on the single common coordinate. [2] However, usually the objects in each group are different and the number of objects in each group is different to each other. Therefore, we cannot use ordinary MDS for 3-way data. In order to solve this problem, we propose a new transformation method based on a cluster-scaling over the different labeled data sets.

First, we create dissimilarity matrices of different labeled data sets as follows:

$$D^{(t)} = \begin{pmatrix} d_{11}^{(t)} & \cdots & d_{1p}^{(t)} \\ \vdots & \ddots & \vdots \\ d_{p1}^{(t)} & \cdots & d_{pp}^{(t)} \end{pmatrix}, \ d_{ab}^{(t)} = \sum_{i=1}^{n_t} (x_{ia}^{(t)} - x_{ib}^{(t)})^2, \ a, b = 1, \cdots, p, \quad t = 1, \cdots, T. \tag{3}$$

Next, we create a dissimilarity matrix of all objects through different labeled data sets as follows:

$$D = \begin{pmatrix} d_{11} & \cdots & d_{1p} \\ \vdots & \ddots & \vdots \\ d_{p1} & \cdots & d_{pp} \end{pmatrix}, \ d_{ab} = \sum_{i=1}^{n} (x_{ia} - x_{ib})^2, \ a, b = 1, \cdots, p. \tag{4}$$

From equations (2), (3), and (4), the following equation is derived:

$$D = \sum_{t=1}^{T} D^{(t)}. \tag{5}$$

In order to obtain a result of MDS on a single coordinate from the different dissimilarity matrices $D^{(t)}, t = 1, \cdots, T$, we must transform $D^{(t)}, t = 1, \cdots, T$ to a single scaled data. As this scale, we use the obtained clusters from the dissimilarity matrix D. Since D does not depend on each labeled data but depends on all of data over the different

labeled data, the obtained clusters from the dissimilarity matrix D can be used for a scale to measure the different dissimilarity matrices on the common bases.

In order to obtain a clustering result from D, we use a fuzzy clustering method named FANNY algorithm. [6] The purpose of this clustering is to classify the p variables into K clusters. The state of fuzzy clustering [1] is represented by a partition matrix:

$$U = (u_{ak}), \quad a = 1, \cdots, p, \ k = 1, \cdots, K, \tag{6}$$

whose elements show the degree of belongingness of the variables to the clusters. In general, u_{ak} satisfies the following conditions:

$$u_{ak} \in [0, 1], \quad \sum_{k=1}^{K} u_{ak} = 1. \tag{7}$$

The objective function of FANNY algorithm is defined as follows:

$$J(U) = \sum_{k=1}^{K} \left(\sum_{a=1}^{p} \sum_{b=1}^{p} u_{ak}^{m} u_{bk}^{m} d_{ab} / (2 \sum_{l=1}^{a} u_{lk}^{m}) \right). \tag{8}$$

The exponent m which determines the degree of fuzziness of the clustering is chosen from $(1, \infty)$ in advance. By minimizing equation (8) under the conditions shown in equation (7), we obtain the solution U.

From equations (5) and (6), we define the following transformation:

$$DU = \sum_{t=1}^{T} D^{(t)} U. \tag{9}$$

In equation (9), (a, k)-th element of DU is as follows:

$$\sum_{b=1}^{p} d_{ab} u_{bk}, \tag{10}$$

and (a, k)-th element of $D^{(t)}U$ is as follows:

$$\sum_{b=1}^{p} d_{ab}^{(t)} u_{bk}. \tag{11}$$

In equation (10), if variables a and b are close to each other, that is d_{ab} is small but a degree of belongingness of a variable b to a cluster k is large, that is u_{bk} is large, then a degree of relationship between the variable a and a cluster k becomes larger. That is the product $d_{ab}u_{bk}$ shows the measured value of a degree of relationship between the variable a and a cluster k through the commonly related variable b with both a variable a and a cluster k. In this sense, DU shows degree of relationship between variables and clusters. In addition, in DU, since d_{ab} is obtained related with all objects, the variables of DU are also related with all objects. However, in $D^{(t)}U$, from equation (11), since $d_{ab}^{(t)}$ is obtained only related with objects which belong to a group t, the variables of $D^{(t)}U$ are also only related with the objects of a group t.

However, in equations (10) and (11), since we use the same clusters over the DU and $D^{(t)}U$, $t = 1, \cdots, T$, the variables in DU and $D^{(t)}U$ are comparable based on the same scaling of the same clusters.

From equation (9), the following equation is obtained.

$$D^{(t_1)}U = DU - (D^{(1)}U + \cdots + D^{(t_1-1)}U + D^{(t_1+1)}U + \cdots + D^{(T)}U), \quad \exists t_1 \in \{1, \cdots, T\}. \tag{12}$$

In equation (12), $D^{(t_1)}U$ shows that the degree of relationship between the clusters and variables explained by objects which belong to a group t_1. And $D^{(1)}U + \cdots + D^{(t_1-1)}U + D^{(t_1+1)}U + \cdots + D^{(T)}U$ shows that the degree of relationship between the clusters and variables explained by objects which belong to other groups excluding a group t_1. Therefore, $D^{(t_1)}U$ can be defined as the remained part of degree of relationship between the clusters and variables excluded the

parts explained by other groups. Therefore, using $\tilde{X}^{(t)} \equiv D^{(t)}U,\ t = 1, \cdots, T$ in equation (12), we create the following $pT \times K$ super matrix $\tilde{X}$ under the same K clusters:

$$\tilde{X} = \begin{pmatrix} \tilde{X}^{(1)} \\ \vdots \\ \tilde{X}^{(T)} \end{pmatrix}. \tag{13}$$

Notice that equation (13) shows that each labeled data $\tilde{X}^{(t)},\ t = 1, \cdots, T$ can be transformed to a single scaled data $\tilde{X}$ by using a single cluster-based scaling.

4. Multidimensional Cluster Scaling (MDCS)

Multidimensional scaling (MDS) is a method for capturing efficient information from observed dissimilarity data by representing the data structure in lower dimensional spatial space. As a metric MDS (principal coordinate analysis), the following model [4],[7] has been proposed.

$$d_{ij} = \{\sum_{\lambda=1}^{R} d^{\kappa}(\hat{x}_{i\lambda}, \hat{x}_{j\lambda})\}^{\frac{1}{\kappa}} + \varepsilon_{ij}. \tag{14}$$

In equation (14), d_{ij} is an observed dissimilarity between objects i and j and $\hat{x}_{i\lambda}$ is a point of an object i with respect to dimension λ in R dimensional configuration space. ε_{ij} is an error. $d^{\kappa}(\hat{x}_{i\lambda}, \hat{x}_{j\lambda})$ shows dissimilarity between objects i and j with respect to dimension λ and usually $d^{\kappa}(\hat{x}_{i\lambda}, \hat{x}_{j\lambda}) = \mid \hat{x}_{i\lambda} - \hat{x}_{j\lambda} \mid^{\kappa}$.

That is, MDS finds R dimensional scaling (coordinate) $(\hat{x}_{i1}, \cdots, \hat{x}_{iR})$ and throws light on the structure of similarity relationship among the objects by representing the observed d_{ij} as the distance between a point $(\hat{x}_{i1}, \cdots, \hat{x}_{iR})$ and a point $(\hat{x}_{j1}, \cdots, \hat{x}_{jR})$ in R dimensional space. In equation (14), we use Euclidean distance when $\kappa = 2$. Since we use Euclidean distance, the results of equation (14) is equivalent to R principal components in the principal component analysis. [5] This is implemented under an assumption of

$$\sum_{i=1}^{n} \hat{x}_{i\lambda} = 0, \quad \forall\lambda,$$

due to the double centering procedure.

Based on the ordinary MDS, we propose MDCS as follows:

$$\tilde{d}_{ij} = \{\sum_{\lambda=1}^{R} d^{\kappa}(\hat{x}_{i\lambda}, \hat{x}_{j\lambda})\}^{\frac{1}{\kappa}} + \varepsilon_{ij},\ i, j = 1, \cdots, pT. \tag{15}$$

In equation (15), $\tilde{d}_{ij}$ is a dissimilarity between i-th and j-th rows in $\tilde{X}$ in equation (13). In addition, the following equation can be derived:

$$\hat{x}_{i\lambda} = \hat{x}^{(t)}_{i_t\lambda},\ \ i_t = i - (t-1)p,\ i_t = 1, \cdots, p,\ t = 1, \cdots, T,\ i = 1, \cdots, pT. \tag{16}$$

In equation (16), $\hat{x}^{(t)}_{i_t\lambda}$ shows a point of a variable i in a labeled data group t with respect to dimension λ in R dimensional configuration space. Since the R dimensional configuration space is a vector space through Euclidean space, for a fixed t, we can obtain the following average vector for each t labeled data in the R dimensional configuration space:

$$\boldsymbol{g}^{(t)} = (\bar{x}^{(t)}_1, \cdots, \bar{x}^{(t)}_R),\ \bar{x}^{(t)}_{\lambda} = \frac{1}{p}\sum_{i_t=1}^{p} \hat{x}^{(t)}_{i_t\lambda}. \tag{17}$$

5. Numerical Example

We use an evaluation data for student collaborative homework assignments completed through a web site with both computer and mobile device access capabilities. Knowledge is built within a community through the social interactions of its peers. [8] Collaborative learning (CL) is a social construction of knowledge where individuals are members of a group, that remains engaged in a shared task in which social interaction between peers is fundamental to achieving learning. [3] In a CL activity, three to five members take part in a coordinated effort to learn a specific educational objective. [3]

The participants were all volunteers between the ages of 18 and 19 years of age which were divided into four permanent groups of 7, 8, 7, and 6 students respectively. Each group was given an identical weekly assignment related to the previous class topic which involved collaborative on-line discussions requiring a consensus answer for completion. The 28 students were asked to submit their homework assignments over 16 times through the web site. Data is collected as log data indicating access time, difference of students, devices, and details of actions over 8 months. The submission has been summarized into two categories of activities which are "post" and "read". Post is when the students post their response to the group activity and read is when they read other student posts. In addition, this data is summarized into the value of the data which shows the frequency of access to the web site with respect to each homework assignment and each kind of activity. The data also has an additional two categories of web site access of mobile phone access and computer access.

Based on this data, the target data in this study is shown in table 1. In this table, "Variables" means each homework assignment with an action. For example, "H1P" means first homework and the action of the student is posting a message to other students in the same group. "H1R" means first homework and the action of the student is reading messages posted from other students in the same group. "S1C" means first student and the tool is computer. "S1M" means first student and the tool is mobile phone. These students are classified into four groups "G1", "G2", "G3", and "G4". Notice that each group includes a different number of students. The value of this table shows the number of messages each student posted or read using a computer or mobile phone for each homework assignment.

The purpose is to capture the relationship over the four groups in lower dimensional space. Using data shown in table 1, we obtain the dissimilarity matrices $D^{(1)}$, $D^{(2)}$, $D^{(3)}$, and $D^{(4)}$ for students included for each group "G1", "G2", "G3", and "G4" using equation (3). In addition, we obtain a dissimilarity matrix D through all students using equation (4). Using the obtained D and FANNY algorithm shown in equation (8), we obtain a result of fuzzy clustering shown in equation (6). Here, we assume the number of clusters is 3 and the value of parameter m is 2 in equation (8). Then using the obtained $D^{(1)}$, $D^{(2)}$, $D^{(3)}$, $D^{(4)}$, and U, we obtain the transformed data $\tilde{X}$ shown in equation (13) using equation (12). Then we apply the data $\tilde{X}$ to the proposed MDCS shown in equation (15) and obtain the result shown in equations (16) and (17).

Figure 1 shows a result of fuzzy clustering shown in equation (6). In this figure, "h1r" means a variable "H1R". Each value shows degree of belongingness of a variable to a cluster. From this figure, it can be seen that homework 1 to 7 for reading action has clear separation from other homework activities. Figure 2 shows a result of points of each variable in each labeled data group with respect to each dimension in two dimensional configuration space shown in equation (16). In this figure, for example, "G1h1r" means a point of variable "H1R" in group "G1". This result was obtained by using the clustering result shown in figure 1 as a cluster-based scaling, in order to be comparable different dissimilarity structure of the four groups "G1", "G2", "G3", and "G4". In fact, from figure 2, we can see that not only homework 1 to 7 for reading activities are separated from other activities, but also we can see that in particular, homework 1 of reading of groups "G1", "G3", and "G4", and homework 4 to 7 of reading of groups "G1", "G2" are separated from other activities. This means that in figure 2, we can see which groups of students contribute to the separation of each of the homework activities. Figure 3 shows a result of points of groups shown in equation (17). Figure 4 shows a result of ordinary MDS. This result is obtained using the average of students of each group with respect to variables using the data shown in table 1 and applying the average to an ordinary MDS shown in equation (14). From the comparison between figures 3 and 4, it can be seen that the difference between the two results is large. In fact, the similarity "G2" and "G3", "G4" and "G1" are large in MDCS, however the similarity "G1" and "G2" are large in MDS. In figure 2, proportion of this result for the first eigen value is 0.98 and the proportion of the summation over the first and second eigen values is almost 1. Therefore, it can be seen that this result is obtained almost perfectly fitted to the data. However, the result shown in figure 4, the proportion of the first eigen value is 0.56 and the proportion of the summation over the first and second eigen values is 0.86. It can be seen that our proposed

MDCS can obtain a more accurate result when compared with the ordinary MDS.

Table 1 Students Mobile Activity Learning Data

Groups	Students	Variables				
		H1P	H1R	·····	H16P	H16R
G1	S1C	0	0	·····	0	0
	S1M	4	5	·····	0	0
	⋮	⋮	⋮	⋮	⋮	⋮
	S7C	2	7	·····	0	0
	S7M	8	4	·····	1	5
G2	S8C	0	0	·····	0	0
	S8M	4	5	·····	0	0
	⋮	⋮	⋮	⋮	⋮	⋮
	S15C	2	7	·····	0	0
	S15M	8	4	·····	1	5
G3	S16C	0	0	·····	0	0
	S16M	4	5	·····	0	0
	⋮	⋮	⋮	⋮	⋮	⋮
	S22C	2	7	·····	0	0
	S22M	8	4	·····	1	5
G4	S23C	0	0	·····	0	0
	S23M	4	5	·····	0	0
	⋮	⋮	⋮	⋮	⋮	⋮
	S28C	2	7	·····	0	0
	S28M	8	4	·····	1	5

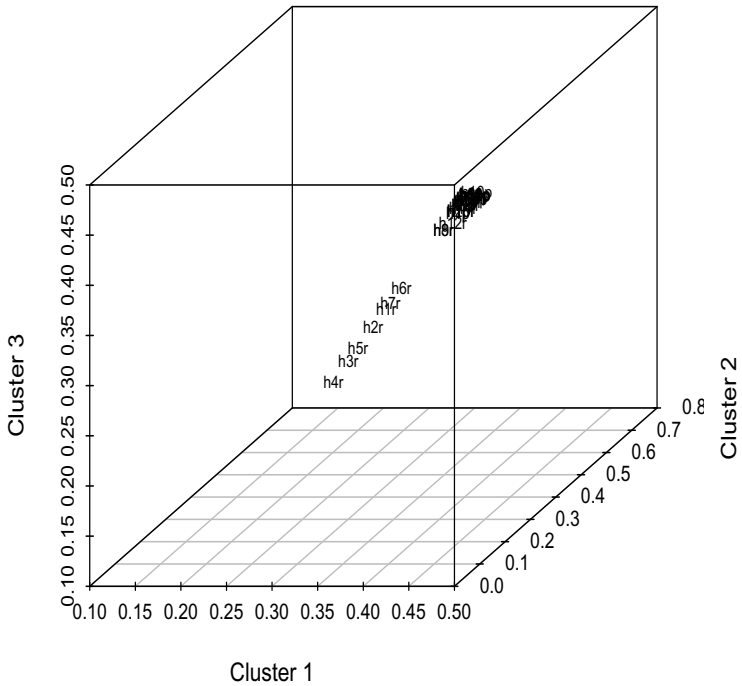

Figure 1 Result of Fuzzy Clustering

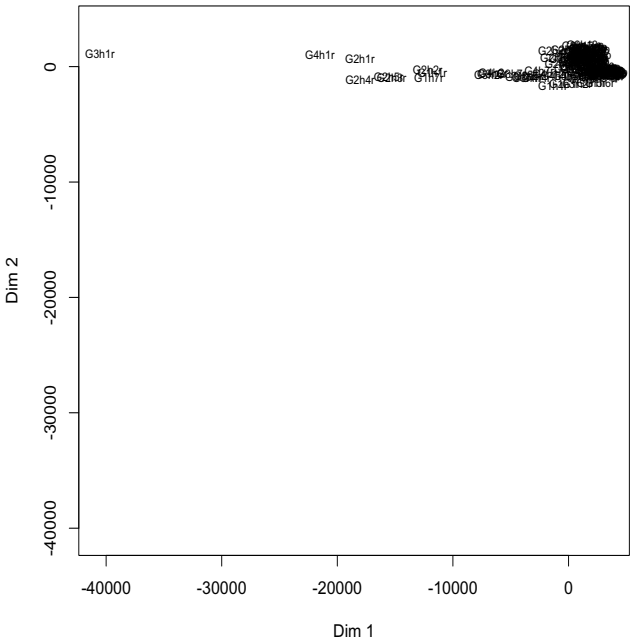

Figure 2 Result of MDCS for Variables

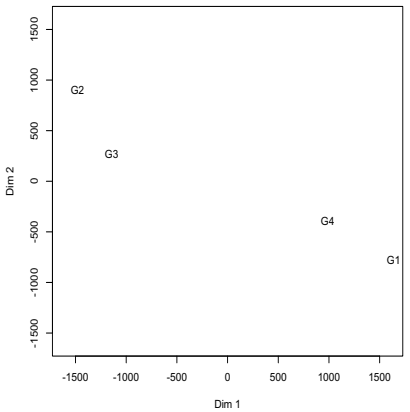

Figure 3 Result of MDCS for Groups

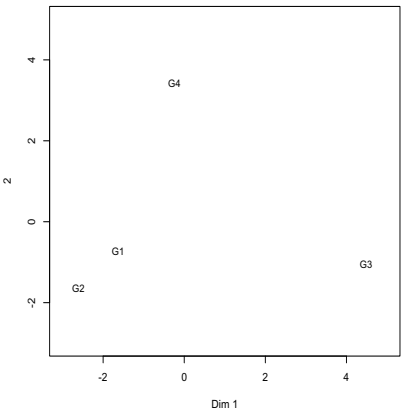

Figure 4 Result of MDS for Groups

6. Conclusion

This paper proposes a multidimensional cluster scaling when we observe a labeled data. Since the multidimensional scaling is based on a coordinate system, we needs a transformation over several labeled data sets based on a single scaling which makes them comparable over the labeled data sets. For this purpose, we use a cluster structure by using the advantage that we can obtain the single classification structure directly from the objects space not based on any coordinate systems. We show a numerical example of mobile learning data and show a better performance by using the proposed method.

[1] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, 1981.
[2] J.D. Carroll and J.J. Chang, Analysis of Individual Differences in Multidimensional Scaling via an N-way Generalization of "Eckart-Young" Decomposition, *Psychometrika*, Vol. 35, No. 3, pp. 283-319, 1970.
[3] P. Dillenbourg, *Collaborative Learning : Cognitive and Computational Approaches*, Amsterdam; Oxford: Pergamon, 1999.
[4] J.C. Gower, Some Distance Properties of Latent Roots and Vector Methods used in Multivariate Analysis, *Biometrika*, Vol. 53, pp. 325-338, 1966.
[5] I.T. Jolliffe, *Principal Component Analysis*, 2nd ed., Springer, 2002.
[6] L. Kaufman and P.J. Rousseeuw, *Finding Groups in Data*, John Wiley & Sons, 1990.
[7] J.B. Kruskal and M. Wish, *Multidimensional Scaling*, Sage Publications, 1978.
[8] L.S. Vygotsky, *Thought and Language*, Cambridge, Mass.: MIT Press, 1986.